DOI: 10.7763/IPEDR. 2012. V49. 22

Research on the Location and Extraction of Texts in Complex Background

Ming Li[1] , Yousheng Zhang[2] and Wei Jia[3]

[1]Special Education Collage, Beijing Union University, Beijing, China

[2]School of Computer Science & Information Technology Anhui Sanlian University, Heifei, China

[3]Intelligent Computing Lab, Hefei Institute of Intelligent , Machines, Chinese Academy of Sciences, Heifei, China

e-mail: li.ming.buu@gmail.com, zhangyos@126.com, icg.jiawei@gmail.com

Abstract—Research on the location and extraction of the texts in complex background has important significance in current information age. It has enriched image processing theory and shows great business value in practical applications such as image and video searching in Internet environment as well as license plate recognition in modern traffic management. How to rapidly and accurately locate and extract the text in image and video with complicate background has become a global hot research topic. This essay take approach from difficult problems encountered by text location and extraction in complicate background in order to analyse text division, feature extraction and recognition processing. The essay also introduces various methods of text location and extraction in complicate background.

Keywords-Complex background; text classification; text feature; location and extraction

1. Introduction

In recent years, with the rapid development of multimedia technology and computer network, the capacity of digital image around the world is increasing at an alarming rate. A capacity of images equal to the number of gigabytes is produced daily. These digital images contain a lot of useful information. Current computer vision and artificial intelligence technology can not automatically mark the image but must rely on manual annotation for image marking, which is not only time consuming but also often inaccurate or incomplete with inevitable subjective bias. That is to say, different people have different understanding methods for the same image, which leads to identification error in image retrieval. This requires a technology that can quickly and accurately search for and access images, which is known as image retrieval. Therefore, how to quickly and accurately locate and extract the text in complex background images and videos has become a hot international research topic.

The complex background refers to (1) the background image(as opposed to other parts of the text) that contains rich texture; (2) sometimes the text is embedded in the texture, and sometimes the text itself is the texture; (3) the possible position, received light, direction, font, size and color of the texts are different, and these information are a priori unknown before location and extraction. These three points are the challenge of this study. If we can find solutions to these problems and construct the model used to locate and extract the text, it would offer important implications to solving other similar complex mode detection problem.

2. Text Feature and Classification

2.1 Text feature

The features included in the text are very rich, as shown

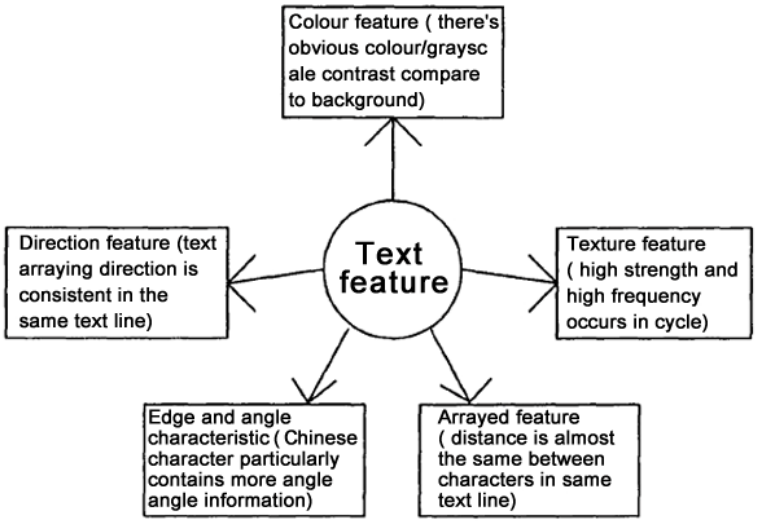

Figure 1. Text feature

in Figure 1. Which of these features are the most useful and how to use these feature is a key question that needs to be studied in terms of text location and extraction.

2.2 Text classification

According to the source, image text is divided into two categories: one is called the scene text and the other is called artificial text.

Scene text refers to the two-dimensional image of scene texts in the real three-dimensional world received by a camera or video camera, such as the license in car pictures, shop signs that occasionally appear on the video screen or texts on street advertisements. Uneven illumination, camera or video camera distortion, lack of exposure and narrow dynamic range, titled shooting angle, uneven surface of three-dimensional texts and various degrees of pollution and other reasons will all lead to low-quality images. Since character size, font, color, direction and background texture of the text is a priori unknown, it is difficult to extract and identify this type of text.

Artificial text is a man-made text, with more standard features and purposes, such as the artificial text in synthesized images and the title sequence, epilogue and subtitles in videos. It can be seen that compared to scene texts in videos, this kind of text has more significant content, which makes it easier to detect and identify and play an important role in video retrieval. Therefore, we mostly use artificial text for the processing of video texts.

In order to have a more detailed understanding of text location and extraction in complex background, Figure 2 shows the difference in image source, color, background complexity and application between scene text and artificial text. (Note: the text location and extraction within a single frame of dynamic images is basically the same with that of static images.

Category / Classification	Scene text	Artificial text
Source of image	Static images (including digitalized pictures, pictures received from digital camera or scanner	Dynamic images(video sequence or film documents), synthesized images on www
Color	Grayness or color	Color
Complexity of background	Spinning, tilted, partially hidden ,partially hidden, uncontrollable light	Both background and text can be mobile
Field of application	Color image database, car license identification	Content-based video search, retrieval and Web search

Figure 2. Text classification

3. Methods of Text Location and Extraction in Complex background

The study of text location and extraction in complex background originates from the extraction of street signs, license plates, shop signs and cargo train signs by Ohya and others[1]. Later, through the identification of cargo containers by Lee, Kankanhalli and others[2], the extraction of CD and book covers by Zhong, Karu, Jain and others and the text extraction of www images by Jiangying Zhou, Daniel Lopresti and others[3], the technology that could only process gray pictures before can no process color pictures and the complexity of the processed background is increasingly high. In recent years, the focus of research is mainly www images[3][4], natural images[1][2][5][6], and videos[7][8][9] (to face the problem of low resolution, in particular, the problem of moving text in the extraction of video text). This new field attracts many researchers, and new algorithms and methods are constantly appearing.

3.1 Connected component-based method

For connection-based text extraction method, it analyzes the geometric arrangement of edges or parts with similar color-grayness, in order to detect text area. Due to the fact that texts have centralized sharp edges, Smith and Kanade[10] define texts as horizontal boxes. Zhong et al[6] extracts texts as connected parts of a single color, and they follow the restrictions on size and horizontal alignment. In a similar manner, Lienhart and Stuber[8] identify text as parts connected in the same color. Within a special range of size, the text has corresponding part in continuous video frames, and the motion estimation of adjacent frame will enhance the effect of text extraction. Jian,Yu[11] separate video frames into different sub-images and then checked if each sub-image contains certain predetermined heuristically searched text. The connected component-based method quickly locate text, but when the text is embedded in complex background or exposed to other text or images (especially when the video texts are embedded in complex background), then using this method will lead to difficulties.

3.2 Texture-based method

Popular texture analysis methods such as Gabor filter[12], Gaussian filter[13] or spatialvariance[6] can be used to locate the text area. Jain and Zhong[12][14] used distinct textures to determine and separate text, graphics and images from scanned gray scale images. In [12], Jain describes using multi-channel based Gabor filters to separate text and image areas. But Le[15] pointed out that Gabor filters are too computationally intensive and cannot perform orthogonal decomposition. Besides, its choice of parameters is related to images, which is not practical. Wu and Manmatha[13] proposed a Gaussian filter-based text extraction system. They view texts as natural textures. After filtering, each pixel in the original image is represented by a feature vector, which is composed of energy calculated from filtered images, and then these feature vectors are combined like in K-means clustering method. Zhong et al[6] further used the texture feature of texts in grayscale images, for each pixel, the text energy is defined as the horizontal spatial difference within a 1*n neighboring window. This texture-based method is applied to many static images. Using texture-based method usually cannot result in accurate text area.

3.3 Edge and corner-based method

An identifiable text area has strong contrast in color or brightness, which is to say that edge / corner information is independent of contrast, font color, font size and other text features. In this way, edge / corner becomes a more reliable feature than font color. Edge / corner has two characteristics: the strength of edge / corner and the density of edge / angle. Edge-based[16][17][18][19] approach breaks down the text area by analyzing the projection of edge strength. Corner-based[20] approach obtains corners image through cornersw detection, then refines and consolidates corners according to the width, size and other features of the text area, obtains candidate text areas and decomposes these areas into single text line using vertical and horizontal edge detection. For Chinese texts, these two approaches are very effective, because Chinese has more and denser edges and corners. However, when the size of the text is very big, the text appears similar to the texture, which has rich corners and edges. When the parameters are adjusted to determine whether the texture with rich edges and corners is non-text, large texts cannot be detected either.

3.4 Artificial neural network-based method

Huiping Lid[21] from University of Maryland used 16 * 16 pixel block-based hybrid wavelet / neural network segmentation method in the detection of video texts. First, they used Haar wavelet decomposition to obtain the texture features of texts and non-texts (texts with very high wavelet coefficients in neighboring areas of edges and corners, making the texts to be detected in high-frequency parts easily). Then they used 16 * 16 window to scan the whole image. Taking into account of speed and accuracy, they scanned every four pixels and used 3-layer BP neural network as classifier to identify texts and non-texts. In order to resolve the shortage of training samples, they used the bootstrap method proposed by Sung and Poggio[22] for the training of a line of samples. Wernicke and Lienhat[17] also adopted bootstrap method in sample training. The classifier used is a two feedforward neural network with the training cycles for classification and identification. They used 20 * 10 window to scan the whole image from left to right and from top to bottom. Using neural network method to classify pixels requires previous training with neutral network for images. For image of relatively large size, although they don't need to be scanned line by line, the training still takes quite a long time, which is not satisfactory for real-time processing.

4. Conclusion

Apart from the above methods, Kwang In Kim [23] first used SVM for text location and extraction in 2000 and achieved achieved good results (Kwang In Kim and others also used SVM in face identification and achieved good results). Jun and others[24] used grayscale image region technology to identify characters in complex scenarios with an identification rate of 71.1%. In addition, a method to extract text in video is given by Lyu and other researchers[25]. They have solved problem of different character size by multi-resolution analysis. They also conducted similar positioning arithmetic disposal for images in different scales after multi-resolution decomposition. In another words, they firstly applied a kind of improved Sobel operator to extract edge, then they used approach of part self-adjustment stop value to transform edge image into binary image, finally projection analysis was applied to conduct positioning of text region. Before conduct projection analysis in candidate text area, Jie Xi and others[26] used mathematical morphological method to binary expand the graph. Liang and other[27] used morphological method to extract text from regular background image without causing any loss to character shape. Tan and others[28] used pyramid method to separate characters from map, which is suitable for the field of GIs. Hwang[29] and others analyzed the reason that characters are affected by noise disturbance in OCR, used wavelet analysis to extract characters and obtained complete characters. Zhou and Lopresti[3] used genetic algorithm to extract texts from grayscale images.

5. Acknowledgment

This research was supportede by the National Science Foundation of China(NSFC) project(60705007).

6. References

[1] J.Ohya, A.Shio, and S.Akamatsu, "Recognizing characters in scene images", IEEE Transactions Pattern Analysis and Machine Intelligence. Los Alamitos USA, vol.16, pp.214-22, Feb 1994.

[2] C-M. Lee and A. Kankanhalli, "Automatic extraction of characters in complex scene images", International Journal of Pattern Rocgnition and Artificial Intelligence. Singapore, vol.9, pp.67-82, 1995.

[3] J.Zhou and D.Lopresti, "Extracting text from WWW images". Proceedings of the Fourth International Conference on Document Analysis and Recognition. Ulm Germany, vol.1, pp.248-252. Aug 1997.

[4] J. Zhou, D.Lopresti, and T.Tasdize, "Finding text in color images". In Prodeedings of SPIE, Document Recognition V. San Jose USA, pp.130-140, January 1998.

[5] B.Yu, A.K.Jain, and M. Modiuddin, "Address block location on complex mail pieces", Proceedings of the Fourth International Conference on Document Analysis and Recognition. Ulm Germany, vol.2, pp.897-901, Aug 1997.

[6] Y.Zhong, K.Karu, and A.K.Jain, "Locating text in complex color images", Pattern ecognition. vol 28, pp.1523-1535, October 1995.

[7] H.Li and D.S.Doermann, "Automatic identification of text in digital video key frames", Fourteenth International Conference on Pattern Recognition, 1998. Proceedings. Brisbane Australia, vol.1 pp.129-132, Aug 1998.

[8] R.Lienhart and F.Stuber, “Automatic text recognition in digital videos”, In proceedings of ACM Multimedia. pp.11-20, 1996.

[9] J.Shin, C.Dorai, and R.Bolle. Automatic text extraction from video for content-based annotation and retrieval. Fourteenth International Conference on Pattern Recognition, 1998. Proceedings. Brisbane Australia, vol.1, pp.618-620, Aug 1998.

[10] M.A.Smith and T.Kanade, “Video Skimming and Characterization throughLanguage and Image Understanding Techniques”, Proceedings 1998 IEEE International Workshop on Content-Based Access of Image and Video Database. Bombay India, pp.61-70, Jan 1998.

[11] A.K.Jain and B.Yu, “Automatic text location in images and video frames”, Pattern Recognition.Vol.31, pp.2055-2076,December 1998.

[12] A.K.Jain and S.Bhattacharjee, “ Text segmentation using Gabor filters for automatic document processing”, Machine Vision and Applications. New York USA, vol.5, pp.169-184,1992.

[13] V.Wu, R.Manmatha, and E.M. Riseman, “Automatic text detection and recognition”, In Proceedings of Image Understanding Workshop. pp.707-712,1997.

[14] A.K.Jain and Y.Zhong, “Page Segmentation Using Texture Analysis”, Pattern Recognition. vol.29, pp.743-770, May 1996.

[15] D.K.Le, G.R.Thoma and H.Wechster, “Classification of binary document images into textural or nontextual data blocks using neural network model”, Machine Vision and Applications. Vol.8,pp.289-304,1995.

[16] W. Qi, et. Al, “Integrating Visual, Audio and Text Analysis for News Video”, 7th IEEE International Conference on Image Processing (ICIP2000). Vancouver British Columbia Canada, pp.10-13, September 2000.

[17] A.Wernicke, R.Lienhart, “On the Segmentation of Text in Videos”, 2000 IEEE International Conference on Multimedia and Expo-ICME 2000. New York USA, vol.3, pp.1511-1514, July-August 2000.

[18] T.Sato, T.Kanade, E.Hughes, and M.Smith, “Video OCR for Digital News Archives”, Proceedings 1998 IEEE International Workshop on Content-Based Access of Image and Video Databases. Bombay India, pp.52-60, January 1998.

[19] XR Chen, HJ Zhang, “Text Area Detection from Video Frames”, Advances in Multimedia Information Processing-PCM 2001. Beijing China, pp.222-228, October 2001.

[20]Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, Hong-Jiangi Zhang, “Automatic Location of Text in Video Frames”, Proceedings of the 2001 ACM workshops on Multimedia: multimedia information retrieval. New York USA, pp.24-27, 2001.

[21] Li. H, Doermann.D, Kia.O, “Automatic text detection and tracking in digital video”, IEEE Transactions on Image Processing. Evanston USA, vol.9, pp.47-156, Jan 2000.

[22]K.Sung and T.Poggio, “Example-based learning for view-based human face detectionm”, IEEE Transactions Pattern Analysis and Machine Intelligence. Los Alamitos USA, vol.20, pp.39-51, Jan 1998.

[23] Kwang In Kim, “Support vector machine-based text detection in digital video”, Neural Networks for Signal Processing X,2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop. Sydney NSW Australial, vol.2, pp.634-641, December 2000.

[24] J.Ohya, A.Shio and S.Akamatsu, “Recognizing Characters in Scene Images”, IEEE Transactions Pattern Analysis and Machine Intelligence. Los Alamitos USA, vol.16, pp.214-220, Feb 1994.

[25] LYUMR, SONG Jiqiang, CA Imin, “A comprehensive method for multilingual video text detection, localization and extraction”, IEEE Transactions on Circuits and Systems for Video Technology. Gaithersburg USA, vol.15, pp.243-255, Feb 2005.

[26] Jie Xi, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, Hong-Jiang Zhang, “A Video Text Detection and Recognition System”, 2001 IEEE International Conference on Multimedia and Expo (ICME2001). Tokyo Japan , pp.1080-1083, August 2001.

[27] S.Liang and M.ahmadi. “A Morphological Approach to Text String Extraction from Regular Periodic Overlapping Text/Background Images”, Image Processing, 1994 Proceedings, ICIP-94, IEEE International Conference. Austin USA, vol.1, pp.144-148, Nov 1994.

[28] C.L.Tan and P.O.NG, “Text Extraction using Pyramid”, Pattern Recognition, vol.31, pp.63-72, January 1998.

[29] Wen L.Hwang and Fu Chang, “Charater Extraction From Document Using Wavelet Maxima”, Image and Vision Computing, vol.16, pp.307-315, 1998